DOCUMENT RESUME

ED 071 047 24 CS 000 306

AUTHOR	Shavelson, Richard J.; And Others
TITLE	The Effects of Position and Type of Question on Learning from Prose: The Interaction of Treatments with Individual Differences in Learners. Research and Development Memorandum No. 98.
INSTITUTION	Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-5-0252
PUB DATE	Dec 72
CONTRACT	OEC-6-10-078
NOTE	23p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Critical Thinking; Educational Research; Individualized Instruction; *Inquiry Training; *Junior College Students; Language Ability; *Questioning Techniques; *Teaching Methods; *Verbal Ability

ABSTRACT

Eighty-seven junior college students were assigned at random to one of four treatment groups for which the position (before or after) and type (lower-order vs. higher-order) of questions inserted in a prose passage were varied, or to a control group for which no questions were inserted in the passage. Students were given five aptitude tests prior to instruction. Instruction consisted of reading a 1525-word prose passage. Students were given an achievement test based on the passage both immediately after instruction and two weeks later. When group means on four measures from the achievement test were compared, the differences were usually small and sometimes statistically significant. The most noteworthy finding of this study was that for students who had scored low on a test of verbal ability, the insertion of higher-order questions facilitated learning such that they performed at levels comparable to students in other treatments who were high in verbal ability. If these findings are replicated, instruction could be individualized by assigning students with low scores on verbal ability to a higher-order question treatment and students with high scores to a question-free treatment. (Author)

FILMED FROM BEST AVAILABLE COPY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED 071041

STANFORD CENTER
FOR RESEARCH AND DEVELOPMENT
IN TEACHING

Research and Development Memorandum No. 98

THE EFFECTS OF POSITION AND TYPE OF QUESTION ON LEARNING FROM PROSE: THE INTERACTION OF TREATMENTS WITH INDIVIDUAL DIFFERENCES IN LEARNERS

Richard J. Shavelson, David C. Berliner,
Michael M. Ravitch, David Loeding

School of Education
Stanford University
Stanford, California

December 1972

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred. (Contract No. OEC-6-10-078, Component 1B.)

CS 000 306

Introductory Statement

The Center's mission is to improve teaching in American schools. Too many teachers still employ a didactic style aimed at filling passive students with facts. The teacher's environment often prevents him from changing his style, and may indeed drive him out of the profession. And the children of the poor typically suffer from the worst teaching.

The Center uses the resources of the behavioral sciences in pursuing its objectives. Drawing primarily upon psychology and sociology, but also upon other behavioral science disciplines, the Center has formulated programs of research, development, demonstration, and dissemination in three areas. Program 1, Teaching Effectiveness, is now developing a Model Teacher Training System that can be used to train both beginning and experienced teachers in effective teaching skills. Program 2, The Environment for Teaching, is developing models of school organization and ways of evaluating teachers that will encourage teachers to become more professional and more committed. Program 3, Teaching Students from Low-Income Areas, is developing materials and procedures for motivating both students and teachers in low-income schools.

This memorandum reports research conducted by the Assessment System component of the Program on Teaching Effectiveness. One facet of the Assessment System deals with the placement of trainees in the Model Teacher Training System. A placement decision can be of two types: (1) does the trainee require training in a particular skill or skills, and (2) if yes, to what training method should the trainee be assigned? The present experiment examined the question of differential assignment of trainees (junior college students, in this study) with regard to differences in the students' aptitudes and differences in training treatments.

Abstract

Eighty-seven junior college Ss were assigned at random to one of four treatment groups for which the position (before or after) and type (lower-order vs. higher-order) of questions inserted in a prose passage were varied, or to a control group for which no questions were inserted in the passage. Ss were given five aptitude tests prior to instruction. Instruction consisted of reading a 1525-word prose passage. Ss were given an achievement test based on the passage both immediately after instruction and two weeks later. When group means on four measures from the achievement test were compared, the differences were usually small and sometimes statistically significant. The most noteworthy finding of this study was that for Ss who had scored low on a test of verbal ability, the insertion of higher-order questions facilitated learning such that they performed at levels comparable to students in other treatments who were high in verbal ability. If these findings are replicated, instruction could be individualized by assigning students with low scores on verbal ability to a higher-order question treatment and students with high scores to a question-free treatment.

THE EFFECTS OF POSITION AND TYPE OF QUESTION ON LEARNING FROM PROSE: THE INTERACTION OF TREATMENTS WITH INDIVIDUAL DIFFERENCES IN LEARNERS

Richard J. Shavelson
Stanford University

David C. Berliner
Far West Laboratory for Educational Research and Development

Michael M. Ravitch
Stanford University

David Loeding
Stanford University

Studies on mathemagenic activities (Rothkopf, 1965; 1970) have focused on intended and incidental learning as a consequence of inserting questions in prose material. The effect of question position (before vs. after prose passages) has been examined thoroughly; questions inserted after prose passages facilitate intended and incidental learning, whereas questions inserted before only facilitate intended learning (Frase, 1968; Rothkopf, 1966; Rothkopf & Bisbicos, 1967). More recently the effects of the type of question (e.g., application, synthesis, or higher-order question vs. factual or lower-order question) have been explored (Allen, 1970; Hunkins, 1968; Tenenberg, 1969; Watt & Anderson, 1971). For example, Hunkins' data indicate that higher-order questions prompt more thorough study and cognitive reorganization of the material, while lower-order questions direct attention to facts. Similarly, Watts and Anderson have found that questions requiring application of the learned material, as contrasted to questions that repeat examples or questions on people's names, facilitate learning by prompting the student to inspect and comprehend the text more thoroughly. Carroll (1971), in a review of this area, speculated that "questions are most effective when they not only cause memory search, but also cause some sort of reorganization of memory traces and associations [p. 164]."

When mean differences between treatment groups are examined, the effects of position and type of question are statistically significant, though usually small. But let us suppose that for certain Ss with similar scores on personality or ability measures, the effects of these variables are quite different from their effects on other Ss in the treatment group; such differential effects cannot be detected when a mean is calculated for all the Ss in the group. A certain combination of treatment variables may facilitate learning for a particular subgroup of Ss and not for others. These are hypotheses about aptitude-treatment interactions, or ATI (see Berliner & Cahen, in press; and Cronbach & Snow, 1969).

Two recently completed ATI studies are related to the present investigation. Berliner (1971) examined the effects of factual questions placed at specified intervals in a lecture presentation. In many analyses, for Ss with low scores in memory ability, questions inserted after

lecture segments facilitated learning as measured by both immediate and delayed short-answer tests. For Ss with high scores in memory ability, the questions may have interfered with learning. In a prose learning study, Hollen (1971) found disordinal interactions with associative memory. When the treatment did not include questions, the need for associative memory was maximized; when the treatment included questions, the need for asociative memory was minimized. To optimize learning, then, Ss with low scores on tests of memory might be assigned to an adjunct-question treatment and subjects with high scores might be assigned to a question-free treatment.

The Berliner and Hollen studies raise questions about the main effects of questions inserted in instructional material. The purpose of this study is to further examine the interactions between a given S's aptitudes and instructional treatments that differ in the position and type of questions inserted in a prose passage.

Method

Subjects

Eighty-seven volunteer Ss from a junior college were placed at random in one of five treatment groups:

LB:	lower-order questions before text	(N = 18)
LA:	lower-order questions after text	(N = 21)
HB:	higher-order questions before text	(N = 20)
HA:	higher-order questions after text	(N = 13)
C:	control, no questions in text	(N = 15)

The proportion of male Ss ranged between .22 in group LB and .45 in groups HB and C. The mean age ranged between 21.3 in group LB and 24.2 in group C.

Materials

The instructional material, a prose passage entitled "The Lisbon Earthquake," described the 1775 Lisbon earthquake and its historical and philosophical ramifications. The material was selected on the basis of novelty, reading level, reading time, reported interest, and corresponding test items. Kropp et al. (1965) had devised test items on "The Lisbon Earthquake" and had had high school students read the text and answer the questions. Their data analyses showed that, in general, the empirical classification of items agreed with an a priori classification based on Bloom (1956). From the empirically determined taxonomic classification of test items, the questions for this study were selected. Lower-order questions required Ss to demonstrate "knowledge": "The size

of the tidal wave which hit the Lisbon harbor area was (1) 30 feet; (2) 40 feet; (3) 50 feet; (4) 60 feet." Higher-order questions required Ss to demonstrate "comprehension," "application," or "analysis" (cf. Bloom, 1956): "A 'mental seismograph' is a (1) scientific device for detecting ideas; (2) figure of speech for the mind; (3) mental record; (4) mechanical device for recording earthquakes."

The 1525-word text was divided into eight sections, the first seven consisting of two paragraphs each and the last section of four short paragraphs. For the four experimental groups, a lower- or higher-order multiple-choice question was inserted either before or after each section of text. Thus, a total of eight questions were inserted in the text for each experimental group. For the LB and HB groups, questions were inserted before each section of text and repeated, with the correct answers, at the end of the section. For the LA and HA groups, questions were placed after each section of text and repeated, with the correct answers, on the following page. The control group read the text without inserted questions and answers.

Questions were assigned to the text and the achievement test in the following way. A pool of 32 multiple-choice questions was created by including two lower-order and two higher-order questions from each of the text's eight sections. For the lower-order question treatment, 8 lower-order questions were randomly assigned to both the text and the achievement test. The other 8 lower-order questions appeared on the achievement test only. The 16 higher-order questions were distributed in the same way. The achievement test, then, used all 32 items. It contained 8 lower-order questions that appeared in the text, 8 higher-order questions that appeared in the text, and 16 questions that were not used in the text.

Instrumentation

Five aptitude measures were administered before the Ss read the passage. Three of these tests were from the battery collected by French, Ekstrom, and Price (1963). The Advanced Vocabulary Test (AV), Part 2, measures verbal comprehension. The Hidden Figures Test (HF), Part 1, measures "the ability to keep one or more definite configurations in mind so as to make identifications in spite of perceptual distractions" (p. 9). The AV and HF tests may be considered tests of different general ability, i.e., intelligence, factors. Since different treatments placed differential emphasis on verbal comprehension and the ability to hold ideas in mind despite distraction, these tests might be expected to interact differently with different treatments. The third test, Letter Span (LS), Part 1, measures "the ability to recall perfectly for immediate reproduction a series of items after only one presentation of the series" (p. 26). This test was chosen to replicate an interaction between the ability measured and questioning treatments (Berliner, 1971; Hollen, 1971).

The fourth test was the Taylor Manifest Anxiety Scale (Taylor, 1951). The test, labeled "Biographical Inventory" (BI) in this study, positions Ss on an anxious to non-anxious continuum. It was included because a pilot study showed that high-anxiety Ss performed better on a learning measure when questions preceded text material, whereas low-anxiety Ss performed better when questions followed text material. Presumably the structure provided by questions that preceded the reading passage attenuated the anxiety of high-anxious Ss and permitted them to concentrate on learning the material.

The fifth test, Memory for Semantic Implications (MSI), was constructed specifically for this study. This test was designed to reflect Guilford's (1967) description, in his structure-of-intellect model, of the ability to remember and transform information presented in written material. Such a test appeared desirable because answering a higher-order question seemed to require the ability to remember and transform the information in the text. This hypothesis corresponds to Carroll's (1971) hypothesis about the effects of cognitive reorganization noted above. The reliability (KR-21) of this specially instructed test was .85.

The achievement test has already been described. Most of the items were taken from Kropp et al. Eight new lower-order questions were devised in order to have equal numbers of questions corresponding to each section in the text. The correlation between scores at immediate and delayed-retention testing, a rough index of test-retest reliability, was .71.

Procedures

The experiment was conducted over three one-hour sessions. In the first session, the experimenter explained the sequence of the study, assured confidentiality of test results, urged the Ss to do their best, and explained that the purpose of the study was "to investigate how people learn, particularly how they learn from written materials." Following the introduction, test packages containing the five aptitude tests were distributed. Ss were instructed to write their name, age, and sex on the package. The testing sequence was: AV (4 minutes); BI (7-1/2 minutes); HF (10 minutes); a rest break (4 minutes); LS (5 minutes); and MSI (7 minutes).

The second session was conducted one week later. Instructional materials corresponding to the LB, LA, HB, HA, or C conditions were distributed randomly to Ss. After studying the materials, Ss took an achievement test (immediate posttest).

The third session was conducted two weeks after the second. The achievement test was administered again (retention testing).

After all data were collected and preliminary analyses made, the experimenter returned to the junior college and discussed the study with interested Ss.

Results and Discussion

Measures for Examining Learning

The achievement test contained 32 items classifiable into four groups: (a) lower-order questions that appeared in the text, (b) higher-order questions that appeared in the text, (c) lower-order questions that did not appear in the text, and (d) higher-order questions that did not appear in the text. In addition to a total score measure, scores on the four groups of questions can be combined in several ways to investigate learning. For example, if scores on lower- and higher-order questions that did not appear in the text ("no-text") are combined, a measure of incidental learning is formed. If scores are calculated separately for lower-order and questions/no-text and higher-order questions/no-text, measures of transfer from type of question in-text to the same type of question are formed for lower- and higher-order question groups, respectively. Finally, if scores on test questions that also appear in the text are calculated, a measure of intentional learning is obtained.

Intercorrelations among these measures are presented in Table 1a, (Higher-Order Groups) and Table 1b (Lower-Order Groups). Most are

TABLE 1a

Intercorrelations among Learning Measures Collected Immediately after Instruction (above Principal Diagonal, N = 33) and Two Weeks after Instruction (below Principal Diagonal, N = 26): Higher-Order Groups

Measures	Total Score	Incidental Score	Transfer Score	Intended Score
Total Score	_____	.90*	.73*	.59*
Incidental Score	.91*	_____	.82*	.33*
Transfer Score	.71*	.88*	_____	.23
Intended Score	.65*	.48*	.24	_____

*$p < .05$

TABLE 1b

Intercorrelations among Learning Measures Collected Immediately after Instruction (above Principal Diagonal, N = 36) and Two Weeks after Instruction (below Principal Diagonal, N = 35): Lower-Order Groups

Measures	Total Score	Incidental Score	Transfer Score	Intended Score
Total Score	_____	.88*	.79*	.19
Incidental Score	.90*	_____	.89*	.02
Transfer Score	.73*	.84*	_____	.06
Intended Score	.51*	.24	.12	_____

*$p < .05$

positive and significantly different from zero. The highest are "part-whole" correlations, in which the items on one measure constitute a portion of another. For example, half the items forming the total-score measure constitute the incidental-learning-score measure. The lowest intercorrelations are between transfer and intentional scores. The items comprising these measures are different. Further, some Ss may be at ceiling on the intentional measure (see Table 2), and this restriction of range may be responsible for the lower correlations. Finally, note the difference in the correlations between intended and incidental learning measures in the two tables. This suggests that for lower-order questions, practice, or the inspection of the text, or both, does not facilitate answering questions on incidental material (Table 1b). Higher-order adjunct questions (Table 1a), by contrast, seem to facilitate learning of incidental material

Effect of Placement and Type of Question on Learning: Comparison of Means

Table 2 presents means and standard deviations for each measure by treatment group and time of testing (immediate and retention). Inspection of the number of Ss within each group and the standard deviations, especially for the total score measure, indicates that the assumption of homogeneity of variance in statistical tests will be violated. Thus, al-

TABLE 2

Scores by Treatment Group on Immediate and Retention Learning Measures

Measure (Maximum Score)	Lower-Order Questions Before Text	Higher-Order Questions Before Text	Lower-Order Questions After Text	Higher-Order Questions After Text	Control
Total Score (32)					
Immediate					
Mean	20.69	23.50	23.00	25.31	23.00
Standard Deviation	3.42	3.86	3.80	2.87	5.80
Number of Ss	16[a]	20	20	13	14
Retention[b]					
Mean	19.59	21.56	20.06	24.30	21.00
Standard Deviation	3.37	4.49	4.06	3.80	2.94
Number of Ss	17[a]	18	16	10	13
Incidental Learning (16)					
Immediate					
Mean	9.50	11.05	9.75	11.31	11.36
Standard Deviation	2.78	2.76	2.53	1.65	2.98
Number of Ss	16	20	20	13	14
Retention[b]					
Mean	9.12	10.11	8.50	11.40	10.69
Standard Deviation	2.32	2.89	2.68	2.32	2.29
Number of Ss	17	18	16	10	13
Transfer (8)					
(a) Lower Order/No-Text					
Immediate					
Mean	5.31	5.85	5.60	6.31	6.21
Standard Deviation	1.66	1.46	1.31	1.11	1.37
Number of Ss	16	20	20	13	14

TABLE 2--Continued

Measure (Maximum Score)	Lower-Order Questions Before Text	Higher-Order Questions Before Text	Lower-Order Questions After Text	Higher-Order Questions After Text	Control	
Retention[b]						
Mean	4.94	5.22	4.69	6.30	5.70	
Standard Deviation	1.20	1.31	1.45	1.06	1.18	
Number of Ss	17	18	16	10	13	
(b) Higher Order/No-Text						
Immediate						
Mean	4.06	5.20	4.15	5.00	5.14	
Standard Deviation	1.44	1.67	1.57	1.41	1.92	
Number of Ss	16	20	20	13	14	
Retention[b]						
Mean	4.18	4.89	3.81	5.00	5.00	
Standard Deviation	1.38	2.00	1.64	1.70	1.47	
Number of Ss	17	18	16	10	13	
Intentional Learning (8)[c]					H-O	L-O
Mean	7.38	7.40	7.15	7.00	5.15	6.50
Standard Deviation	.72	1.00	1.00	.82	2.03	1.61
Number of Ss	16	20	20	13	14	14
Mean	5.59	6.31	6.28	6.80	5.08	5.23
Standard Deviation	1.12	1.02	1.18	1.23	1.70	1.24
Number of Ss	17	18	16	10	13	13

[a]Because some Ss marked their answer sheets incorrectly, a few cases were lost in immediate testing. These cases were randomly distributed across treatments. Because some of the answer sheets for these Ss were correctly marked in the retention test, there are more Ss in retention testing than in immediate testing.

[b]Two weeks following instructions.

[c]For the four question groups, intended learning is calculated from scores on questions inserted in text. For example, the score for the LB group is calculated from scores on low-order items appearing in text. Since the control group did not receive questions, scores on both high- and low-order questions are reported as a baseline for comparison.

though all the tests are performed at $\alpha = .05$, the exact level of significance cannot be specified. Therefore, the tests of significance reported below should be interpreted with caution. In these analyses, moreover, all comparisons between means are more conservative than they would be if sample sizes were the same for each treatment.

Total score measure. Total score data were examined with a Treatment Group by Time of Testing (5 x 2) analysis of variance (ANOVA) with repeated measures on the Time factor. Means and standard deviations for the comparison are presented in Table 2. The Group effect was significant ($F = 3.37$, $df = 4/65$). Post-hoc comparisons of group means using the Newman-Keuls method for unequal Ns showed that the HA group scored significantly higher than the LB group. Other comparisons of pairs showed no reliable differences between groups. The Time effect, as anticipated, was significant ($F = 28.48$, $df = 1/65$). Scores were higher at immediate testing ($\bar{X} = 23.04$) than at retention testing ($\bar{X} = 21.06$). The Group x Time interaction was not significant ($F < 1.0$). Orthogonal contrasts were performed to test the effects of (a) type of question, (b) position of question in text, and (c) their interaction. Reliable differences were not found. Nevertheless, the means as presented in Table 2 show a generally facilitative effect for higher-order questions as opposed to lower-order questions, and for questions after rather than before the text. These data, though not significant, conform to the findings in the literature reviewed above.

Incidental learning measure. Incidental learning scores were examined with a Treatment Group by Time of Testing (5 x 2) ANOVA with repeated measures on the Time factor. Means and standard deviations for these comparisons are presented in Table 2. The Group effect was significant ($F = 3.00$, $df = 4/65$). Comparisons of pairs showed no reliable differences between group means. The Time effect was significant ($F = 13.49$, $df = 4/65$). Scores were higher at immediate testing ($\bar{X} = 10.53$) than at retention testing ($\bar{X} = 9.81$). The group x Time interaction was not significant ($F < 1.0$). Orthogonal contrasts were made to test the effects of (a) type of question, (b) position of question, and (c) their interaction. Reliable differences were not found. The trends among the means in this analysis conform to the findings on question position and type.

Transfer learning measure. The scores of groups LB, LA, and C on lower-order test items that were not included in the text can be compared. Similarly, the scores of groups HB, HA, and C on higher-order test items that were not included in the text can be compared. In both cases, these scores can be interpreted as transfer measures. Experience with higher-order questions in the text, say, might facilitate answering higher-order questions that had never been seen before. Two separate Treatment Group by Time of Testing (3 x 2) ANOVAs with repeated measures on the Time factor were performed on the transfer measures for lower-order groups plus the control group and higher-order groups plus the control group. Means and standard deviations for these comparisons are presented in Table 2. For the comparison among the lower-order question groups and control group,

the Group main effect was not significant ($F = 2.11$, $df = 2/40$). The Time main effect was again significant ($F = 8.66$ with 1/40 df). Subjects scored higher at immediate testing ($\bar{X} = 6.66$) than at retention testing ($\bar{X} = 5.69$). The Group x Time interaction was not significant ($F < 1.0$). For the comparison among the higher-order question groups and control group no significant differences were found: Group ($F < 1.0$), Time ($F = 1.84$, $df = 1/36$), and Group x Time ($F < 1.0$). Apparently, then, a transfer effect, from a certain type of question placed in text to the same type of question never before seen, does not occur.

Intentional learning measure. The scores of Groups LB, LA, and C on the lower-order questions that appeared in the text and on the achievement test can be compared. Similarly, the scores of groups HB, HA, and C can be compared on the higher-order questions that appeared in the text and on the achievement test. Thus two separate Treatment Group by Time of Testing (3 x 2) ANOVAs with repeated measures on the Time factor were performed on the intentional scores for lower-order groups plus the control group and higher-order groups plus the control group, respectively. Means and standard deviations for these comparisons are presented in Table 2. For the comparison among the lower-order question groups and control group, a significant Group main effect was obtained ($F = 3.52$, $df = 2/40$); LB and LA groups perform significantly better than the control group. The Time main effect was significant ($F = 19.30$, $df = 1/40$). Subjects scored higher at immediate testing ($\bar{X} = 6.62$) than at retention testing ($\bar{X} = 5.71$). The Group x Time interaction was not significant ($F = 1.72$, 2/40 df). For the comparison among higher-order question groups and control group, a significant Group main effect was obtained ($F = 6.05$, $df = 2/36$); HB and HA groups perform significantly better than the control group. The Time main effect was not significant ($F = 1.32$, $df = 1/36$), nor was the Group x Time interaction ($F = 1.06$, $df = 2/36$). As expected, exposure to questions in the text produces a performance near ceiling on the same items on the achievement test.

In the discussion of the results from the four learning measures, the performance of the control group cannot be overlooked. Except for the HA group, the control group usually performed as well as the experimental groups or better. Although the explanation for this finding is not clear, it may reflect the brevity of the prose material.

Aptitude-treatment interactions (ATI). If the five treatment conditions have similar effects on all Ss and if achievement test scores are regressed on aptitude scores, the regression slopes for treatments should be parallel and the difference among slopes can be explained by the differences in means between the groups. But if the treatment groups do not have similar effects on all Ss (say persons high on aptitude A do well with higher-order questions placed after text, but not with these questions placed before text), the regression slopes for treatments should not be parallel. Rather, they should interact and perhaps cross at some point. To determine if ATI's were present, total and incidental learning scores were regressed on aptitude scores. Transfer and intentional-learning measures were not examined because of the restricted range (0-8 points) and ceiling effect for these data.

TABLE 3

Means and Standard Deviations for Aptitude Variables

		LB	LA	HB	HA	C
AV:	Mean	7.22	9.38	8.90	8.54	9.33
	S.D.	(1.80)	(3.56)	(2.57)	(3.28)	(4.10)
BI:	Mean	16.39	12.62	13.60	12.54	14.73
	S.D.	(8.69)	(6.73)	(7.29)	(6.30)	(5.91)
HF:	Mean	4.06	4.62	4.90	3.77	6.13
	S.D.	(2.92)	(3.89)	(2.36)	(2.92)	(4.50)
LS:	Mean	4.78	4.95	4.95	5.54	4.60
	S.D.	(1.50)	(1.72)	(1.43)	(2.15)	(1.40)
MSI:	Mean	15.61	14.43	14.55	15.15	14.20
	S.D.	(3.00)	(3.84)	(2.78)	(4.30)	(4.59)

Means and standard deviations for each aptitude test are presented separately for each treatment group in Table 3. The data appear to be homogeneous across treatment groups. Correlations among aptitude test scores for all Ss combined are presented in Table 4. In general these

TABLE 4

Intercorrelations Among Aptitude Tests

	AV	BI	HF	LS	MSI
Advanced Vocabulary	1.00				
Biographical Inventory	-.22*	1.00			
Hidden Figures	.09	.01	1.00		
Letter Span	.13	-.04	.11	1.00	
Memory for Semantic Implications	.04	.01	.23*	.07	1.00

*$p < .05$; $N = 87$

correlations show the five aptitude measures to be independent. The correlation between AV and BI and also the correlation between HF and MSI are the exceptions. In either case, only five percent of the variance in one test can be predicted from the other.

The Johnson-Neyman (1936) technique with the Potthoff (1964) modification was used to examine the data for interactions. A computer program for this technique was developed by Dowaliby and Berliner (1971). For each set of analyses, the hypothesis of a common slope (achievement regressed on aptitude) for every possible pairing of treatment groups is tested ($\alpha = .05$). For significant interactions, the Johnson-Neyman technique is applied in such a way that a region of non-significance is determined ($\alpha = .10$). At this point, the alpha level should be determined by the decision-maker since it depends on the type of risk he is willing to take in classifying students. Cases falling within the region of non-significance may be assigned to either treatment; cases falling outside this region should be assigned to one or another treatment.

This analysis did not support the hypothesized interaction between the memory ability tests (LS and MSI) and treatments. Also, the BI and HF tests did not interact with treatments.[1]

Total score measure. Advanced Vocabulary test scores interacted with treatments at immediate and retention testing (Figures 1a and 1b). A disordinal interaction appeared with immediate testing data. The regression slope for the HA group differs significanctly from the slope for the LA and C groups. Ss in the LA and HA groups with AV scores above 8.60 may be assigned to either treatment. Ss with scores below 8.60 on AV should be assigned to the HA treatment. Of the 33 Ss in the LA and HA groups, 19 or 56 percent should be assigned to the HA groups to optimize their achievement. Those subjects in the HA and C group with AV scores above 8.09 may be assigned to either treatment. Subjects with scores below 8.09 should be assigned to the HA treatment. Of the 25 subjects in the HA and C groups, 18 or 67 percent should be assigned to the HA treatment.

For retention testing data, the regression slope for the HA group differs significantly from the slope of the LA group. Ss with scores above 8.98 on AV may be assigned to either treatment. Ss with scores below 8.98 should be assigned to the HA treatment. Of the 26 Ss in the HA and LA groups, 16 or 57 percent should be assigned to the HA treatment.

Incidental learning measure. An interaction with Advanced Vocabulary and treatments using the incidental learning measure was also found. On the basis of immediate testing, Ss in LA and HA groups with aptitude scores below 5.27 should be assigned to the HA treatment; Ss with scores above this value may be assigned to either treatment (Figure 2a). Of the 33 Ss in the LA and HA groups, 3 or 9 percent should be assigned to the HA treatment. Ss in HA and C groups with AV scores below 3.37 should

---

[1] ATI data on these measures are available from the first author.

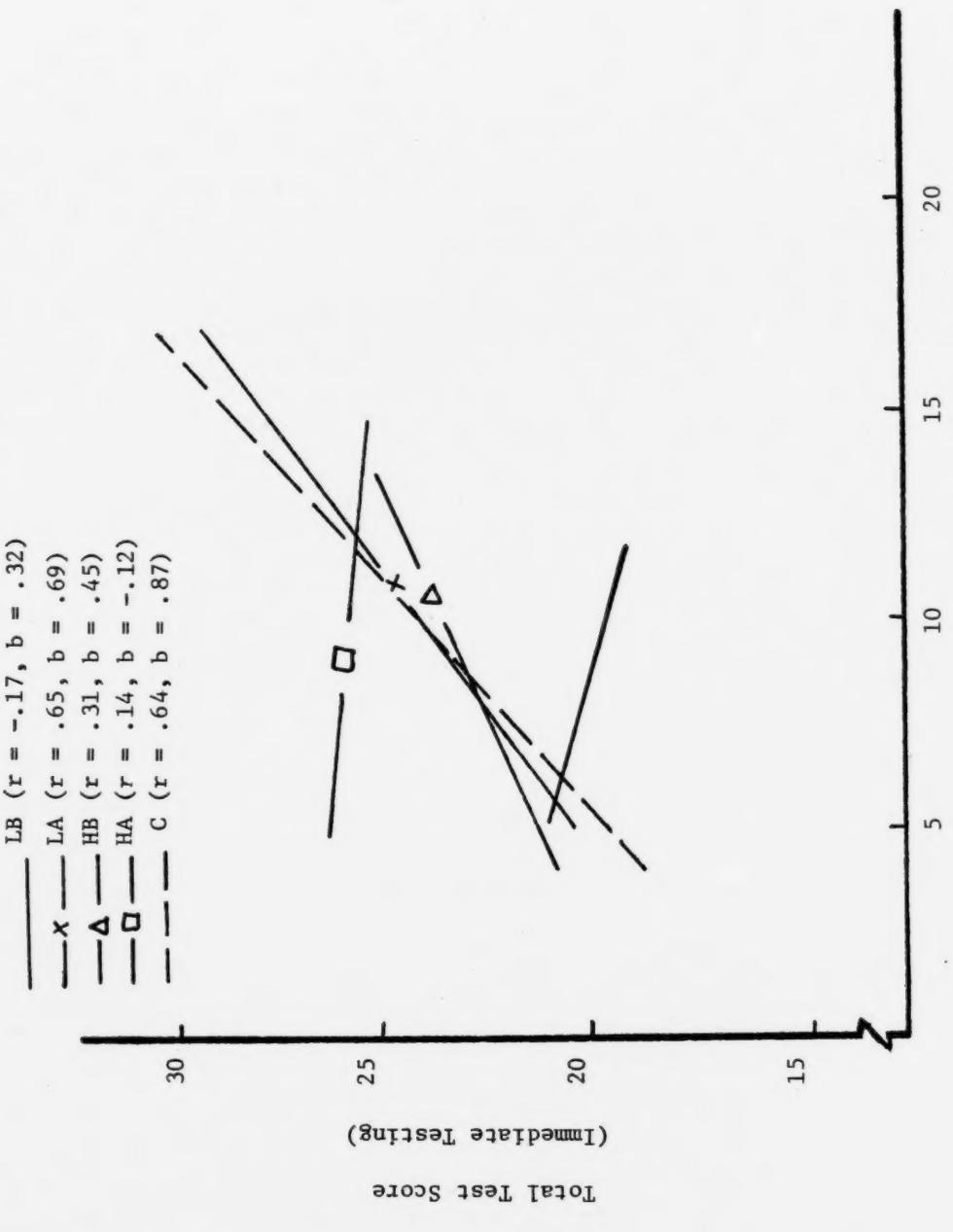

Fig. 1a. Interaction of Advanced Vocabulary scores with treatment effects as measured by the total achievement score at immediate testing.

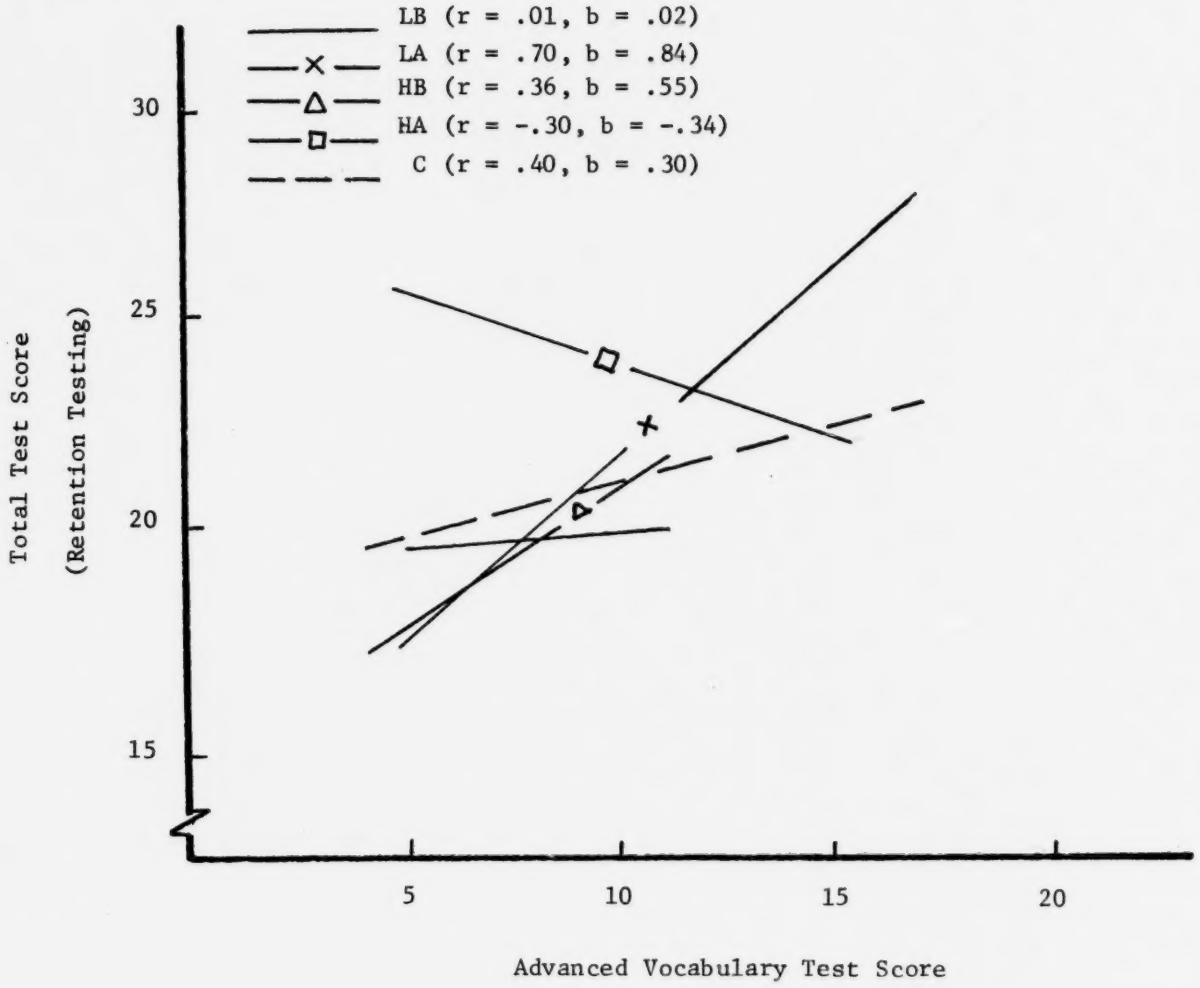

Fig. 1b. Interaction of Advanced Vocabulary scores with treatment effects as measured by the total achievement score at retention testing.

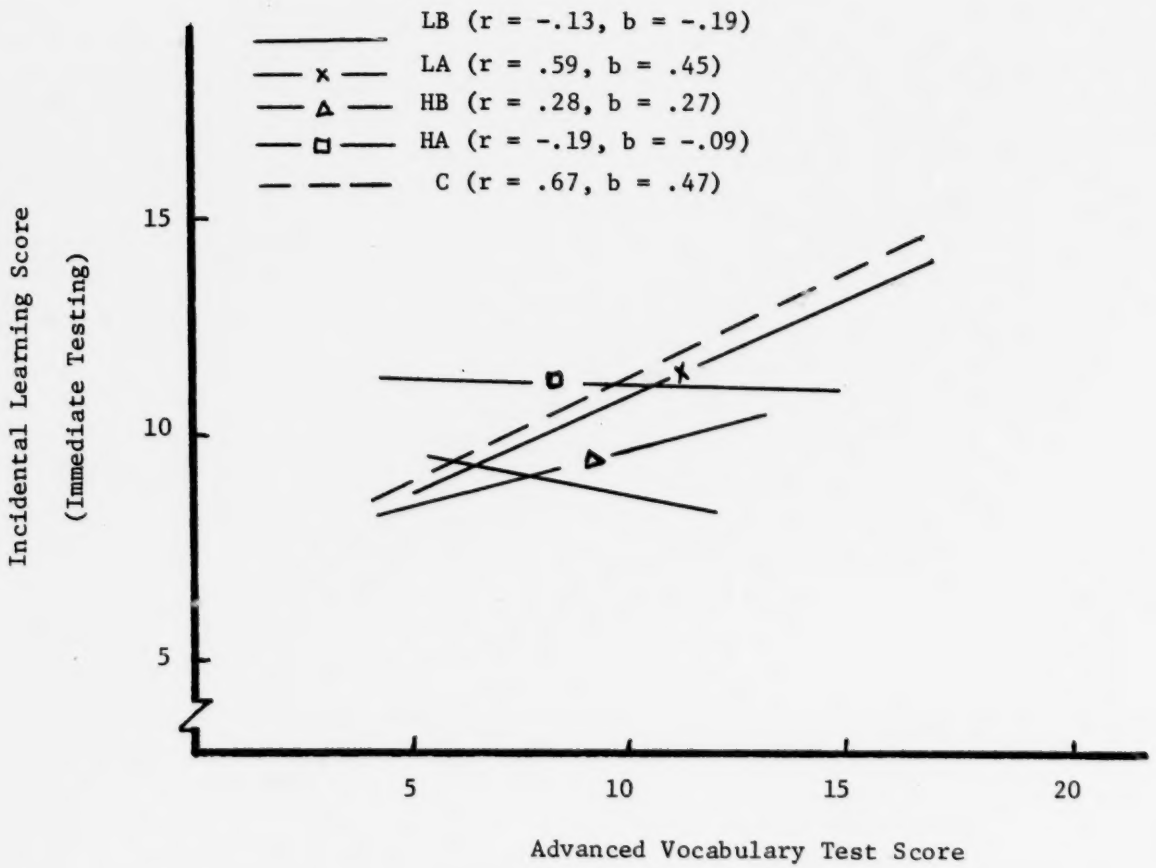

Fig. 2a. Interaction of Advanced Vocabulary scores with treatment effects as measured by the incidental learning score at immediate testing.

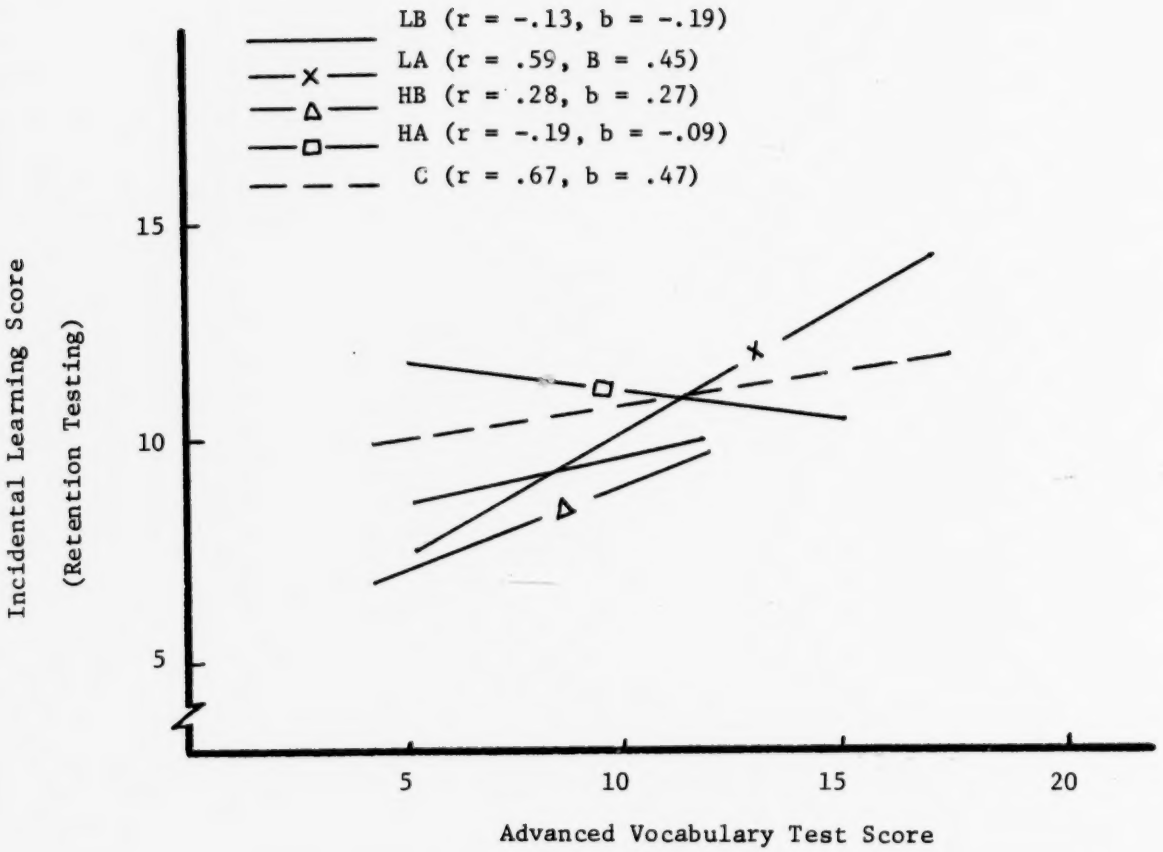

Fig. 2b. Interaction of Advanced Vocabulary scores with treatment effects as measured by the incidental learning score at retention testing.

be assigned to HA; Ss with scores above 17.63 should be assigned to the C treatment. Of the 27 Ss in these groups, none had scores beyond the critical values.

On the basis of retention testing, Ss in the LA and HA groups with AV scores below 8.50 should be assigned to HA; Ss with scores above that value may be assigned to either treatment (Figure 2b). Of the 28 Ss in the LA and HA groups, 16 or 57 percent should be assigned to the HA treatment.

The AV test is related to measures of general mental ability that, according to most studies, yield consistent positive correlations with outcome measures. The unique finding in this study is a slightly negative correlation between scores on the AV test and the outcome measure for Ss in the HA treatment. In conjunction with positive correlations for the LA and C groups, disordinal interactions were obtained.

The insertion of HA questions appears to aid Ss with low AV scores. Such questions may compensate (Salomon, 1971) for deficiencies in the Ss' ability to relate ideas presented in a reading passage. The HA questions may prompt Ss to link concepts in the passage to one another, to the Ss' existing cognitive structure, or to both. The HA question may also stimulate review on the part of Ss who ordinarily would not undertake such an activity.

If replication confirms this finding, immediate steps can be taken to tailor instruction to the needs of particular students. Verbal ability measures are readily available to most teachers. Instructional material for students with low scores on such measures can be modified to include higher-order questions.

Because the negative slope indicates that the HA treatment interferes with learning for Ss high in verbal comprehension, their instructional materials should not include adjunct questions of this nature. These Ss may possess effective strategies for assimilating reading matter that are disrupted by the intrusion of this type of external prompt.

Conclusions

When group means on four measures from the achievement test were compared, the differences were usually small and sometimes statistically significant. Though the effects of position and type of question were not significant, the data agreed with other research findings. Questions following a reading passage proved superior to questions preceding it, and higher-order questions superior to lower-order questions. In the interpretation of these findings, the performance of the control group should be considered. Except for the HA group, the control group usually performed as well as the other experimental groups or better.

Only the interaction between verbal comprehension and treatments was significant. In general, disordinal interactions between HA and C and between HA and LA groups were found in immediate and retention testing for total and incidental scores. The slopes for the regression of achievement scores on AV scores for the C and LA groups were positive. The regression slope for the HA group was negative. The data were interpreted using a compensatory model of ATI.

If replicated, the findings have immediate application. Students with low scores on measures of verbal ability should be assigned reading materials with higher-order questions inserted after the relevant passages. Students with high verbal ability scores should be assigned reading materials that do not include inserted questions.

References

Allen, D. I. Some effects of advanced organizers and level of question on the learning and retention of written social studies material. Journal of Educational Psychology, 1970, 61, 333-39.

Berliner, D. C. Aptitude-treatment interactions in two studies of learning from lecture instruction. Paper presented at the meeting of the American Educational Research Association, New York, N. Y., February 1971.

Berliner, D. C., & Cahen, L. S. Trait-treatment interactions and learning. In F. N. Kerlinger (Ed.), Review of Research in Education. Itasca, Ill.: F. E. Peacock Publishers, in press.

Bloom, B. S. (Ed.) Taxonomy of educational objectives, Handbook I: Cognitive domain. New York: McKay, 1956.

Carroll, J. B. Learning from verbal discourse in educational media: A review of the literature. Research Bulletin No. RB-71-61. Princeton, N. J.: Educational Testing Service, 1971.

Cronbach, L. J., & Snow, R. E. Individual differences in learning ability as a function of instructional variables. Final report, Contract No. OEC-4-6-061269-1217. Stanford University, 1969.

Dowaliby, F. J., & Berliner, D. C. ANALTI: A computer program for the application of the Johnson-Neyman technique in the case of one predictor and one criterion variable. Amherst, Mass. Technical Report No. 15, Center for Educational Research, School of Education, University of Massachusetts, November 1971.

Frase, L. T. Effect of question location, pacing, and mode upon retention of prose material. Journal of Educational Psychology, 1968, 59, 244-49.

French, J. W., Ekstrom, R. B., & Price, L. A. Manual for kit of reference tests for cognitive factors. Princeton, N. J.: Educational Testing Service, 1963.

Guilford, J. P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Hollen, T. T., Jr. Interaction of individual abilities with the presence and position of adjunct questions in learning from prose materials. Dissertation Abstracts International, 1971, 31A, 5847.

Hunkins, F. P. The effects of analyis and evaluation questions on various levels of achievement. Paper presented at the meeting of the American Educational Research Association, Chicago, February 1968.

Johnson, P. O., & Neyman, J. Tests of certain linear hypotheses and their application to some educational problems. Statistical Research Memoirs, 1936, 1, 57-93.

Kropp, R. P., Stoker, H. W., & Bashaw, W. L. The construction and validation of tests of the cognitive processes as described in the "Taxonomy of Educational Objectives." Contract No. OE-4-10-019. Florida State University, 1965.

Potthoff, R. F. On the Johnson-Neyman technique and some extensions thereof. Psychometrika, 1964, 241-56.

Rothkopf, E. Z. Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.), Learning and the educational process. Chicago: Rand McNally, 1965. Pp. 193-221.

Rothkopf, E. Z. Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 1966, 3, 241-49.

Rothkopf, E. Z. The concept of mathemagenic activities. Review of Educational Research, 1970, 40, 325-36.

Rothkopf, E. Z., & Bisbicos, E. E. Selective facilitative effects of interspersed questions on learning from written materials. Journal of Educational Psychology, 1967, 58, 56-61.

Salomon, G. Heuristics for the generation of aptitude treatment interaction hypotheses. Paper presented at the meeting of the American Education Research Association, New York, N. Y., February 1971.

Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. Journal of Experimental Psychology, 1951, 41, 81-92.

Tenenberg, M. S. Effects of different study instructions on learning from written material. Unpublished doctoral dissertation, University of California at Berkeley, 1969.

Watts, G. H., & Anderson, R. C. Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 1971, 62, 387-94.